Personal Identification by Gait Analysis (August 2012)

Alexandre Gadiou

Abstract—**Human recognition at a distance is important in security systems. In this document, is described an efficient identification method which is directly inspired from [3]. The problems with the existing algorithms are the accuracy of motion extraction and computation time. The solution proposed here is to identify people by measuring the closest similarity between 2 distance signals. The results obtained with this method are very encouraging. The accuracy of this algorithm is 95% in outdoor conditions with a database containing 20 people and has a low computational cost.**

Index Terms— **video surveillance, human motion analysis, principal component analysis, gait recognition**

I. Introduction

Biometrics is a technology which is very used to identify people by their characteristics (fingerprints, face, iris…), especially in control access and also in the video surveillance.

Human motion analysis enables to understand the behavior of persons. This process can determine how individuals move (walking or running) for example. In video surveillance, the association of biometrics and human motion analysis is an active research area.

The human identification at a distance has been developed due to the need of person identification systems in sensitive places such as airports, banks... This new research domain aims to distinguish people by the way they walk. The advantage of this new technique is that we don't need to have a contact with the person like in fingerprint recognition. Although, face recognition does not need a contact, it can't be applied at a great distance. Gait analysis is the only way to identify people at a great distance [3].

However, is gait a reliable feature for identification? Some studies [4] [5] have shown that each individual has its own way of walking. Moreover, you can notice that you do identification by gait every day. Indeed, when you recognize someone who is walking at a distance, you analyze his appearance and the dynamics of his walking motion to identify him [2].

The recognition of people depends on the variation of their silhouette when they walk. In this paper, gait motion is supposed to be a succession of static body postures. We expect that the data contained in silhouettes are statistically sufficient to enable the identification. The recognition of a gait is performed by counting the number of silhouettes which correspond to a sequence of postures representing a person registered in a database. The comparison between silhouettes is performed using PCA (Principal Component Analysis). This technique consisting in analyzing data in another space is very accurate. It has already proved its effectiveness in face recognition where PCA is very used [3].

The solution described in this document is a silhouette-based analysis directly inspired from [3]. The difference with [3] is that the identification is performed by matching silhouettes and not a sequence of successive silhouettes. The results obtained with the algorithm presented in this paper are very encouraging. The accuracy is 95% in outdoor conditions with a database of 20 subjects. In addition to that, the computational cost is low.

The rest of the document is spitted into 3 parts. In the first, is described the existing gait recognition methods. The technical approach is presented in the next section. The experimental results are analyzed in the last part of this report.

II. Existing Gait recognition Methods

Although gait recognition is a new biometric technique, there are already gait analysis methods which have been developed [1] [3] [4] [6]. Motion analysis can be classified into two categories: model-based methods and silhouette-based approaches.

Model-based techniques [1] model the human body and try to match this model with each frame of the video. The advantage of these methods is that the movement is extracted robustly on the model [1]. However, the computational cost is high [1].

Silhouette-based approaches [6] can be divided into two classes:

Methods in the first category consider that gait is a succession of static postures. They try to measure the similarity between two sequences of successive silhouettes to perform recognition. The problem with these systems is the variation of the walking speed. If the system learns a walking motion for one person at a speed s_1, the system cannot recognise the person at a speed $s_2 < s_1$.

The second class methods are spatiotemporal analysis and assume that gait is continuous. These techniques measure the similarity between two sequences by aligning them temporally. However, this type of method consumes a significant computation time which is not good in video surveillance [3].

In comparison with silhouette-based methods, our algorithm is less sensitive to the walking speed because we compare silhouettes and not a sequence. The computational cost of our algorithm is cheaper than a spatiotemporal analysis.

III. Technical Approach

A. Overview of System Architecture

The general architecture of the proposed algorithm is illustrated in the figure n°1. It starts by retrieving the video data from a fixed camera. Then, it estimates the current background and subtracts it from each frame. The result of this operation gives the moving objects in the video. Here, we consider that the only moving object in the scene is the human. After this step, the algorithm binarizes this result to get the silhouette of the person. Next, it converts the 2D silhouette into a 1D signal in order to reduce the computational cost. As a result, each individual is represented in the system by a sequence of signals. Then, the system generates a new space from all the signals of each individual in the database. This training operation enables to perform principal component analysis. The recognition is achieved by doing a similarity calculation in this new space for each silhouette extracted in the video. The identified person is the one having the highest number of corresponding silhouettes in the test sequence [3].

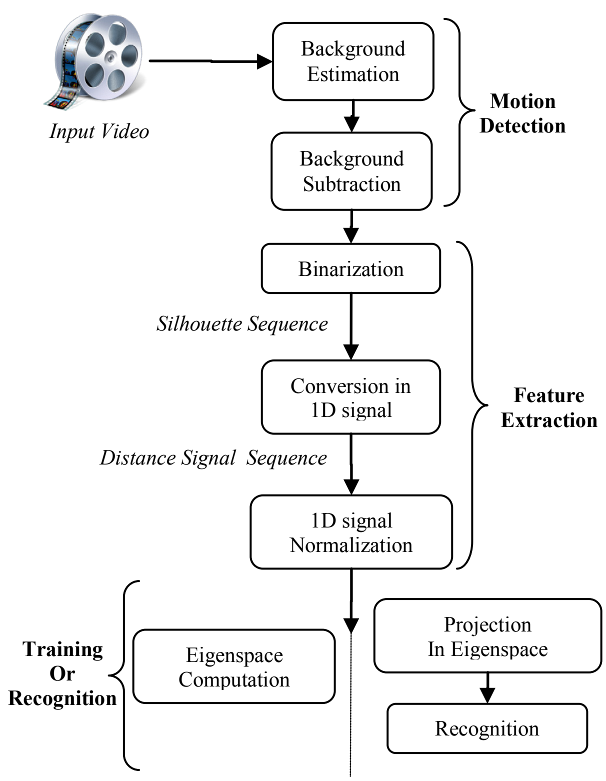


Fig. 1. General architecture of the human recognition system [3]

B. Motion Detection

There are many ways to segment moving object in a video and the most popular method is currently the background subtraction. The fact the camera is fixed, it is easy to extract the moving objects in the scene by subtracting the background. It is impossible to know the real background when a person is walking in the scene. The only way is to estimate the background.

The first approach could have been to take a picture of the scene when nobody is inside. This solution works only if the background is constant which is not true in outdoor conditions. The luminosity is not the same in the morning than in the afternoon. If the morning background is subtracted from an afternoon video sequence, it will be hard to extract the moving objects. In the following paragraphs are described two different ways to estimate the current background in a video.

1) Adaptive Background Estimation

This simple approach makes the assumption that the closest background on the time line is the best estimation of the current background. The expression (1) enables to know when there is motion in a video:

$$D(t) = \frac{1}{N} \sum_{\substack{for \\ each\ pixel}} |I(t_2) - I(t_1)| \qquad (1)$$

I (t_1) and I (t_2) are two successive frames of N pixels. If there is no motion between t_1 and t_2, the result of D will be close to zero because the two frames are the same [7]. The background is easily determined when there is no motion. When there is no motion over 30 successive frames, the background can be defined as the 15^{th} frame. The problem is if the 15^{th} frame is affected by a small change (illumination or small moving object), the accuracy of the background estimation is affected severely. That is why this technique is risky in outdoor conditions.

2) Median Background Estimation

The test of the adaptive background estimation being not satisfactory, another solution based on the definition of the median has been applied. When no motion is detected, the value of pixels is in average constant. However, if there is motion, the value of pixels changes. The value of one pixel has been plot over the time in figure n°2. As we can see in figure n°2, the median seems to be a good approximation of the background. In experiment, this approximation works well if the motion duration is short compared to the size of the video [8].

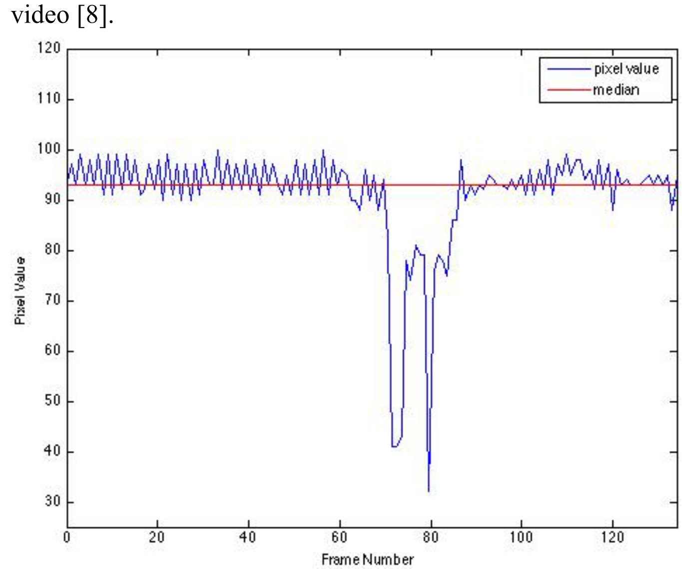


Fig. 2. Value of a pixel over the time

3) *Background Estimation in the System*

A combination of the two background estimations presented previously seems to be the best solution. Indeed, if the background is estimated with median technique and updated every 60 frames, the estimation is less sensitive to slow illumination changes during a day. Moreover, the extraction of the silhouette using background subtraction will be less sensitive to the shadow of the moving person .

4) *Background Subtraction*

This step aims to get the moving object in the video. In order to accentuate the visibility of difference between each frame I and the background B, the algorithm does two operations for each frame of the video.

$$S = \frac{|I(x,y) - B(x,y)|}{B(x,y)} \qquad (2) \qquad [8]$$

$$f(a,b) = 1 - \frac{2 \cdot \sqrt{A_1 B_1}}{A_1 + B_1} \times \frac{2 \cdot \sqrt{A_2 B_2}}{A_2 + B_2} \qquad (3)$$

$$A_1 = (a + \qquad B_1 = (b +$$

$$A_2 = (256 - \qquad B_2 = (256 -$$

a and b represents respectively a pixel value of the frame and a pixel value of the background [3].

C. Feature Extraction

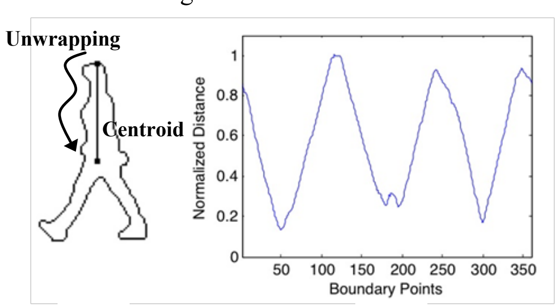
1) *Binarization*

This operation enables to get a black and white image corresponding to the silhouette of the moving person in the video. The program, after having applied (2) and (3), has two results which will be combined after two thresholding operations. A first threshold t_1=0,3 is applied on the result of (2). If $S > t_1$, pixels are considered to belong to the moving object otherwise they are interpreted to be in the background. Another threshold t_2=0.01 is applied at the output of (3).

The combination of these two results enables to get a better silhouette.

2) *Conversion into 1D signal*

This step aims to reduce the dimension of the input data in order to make the algorithm faster.



(a) (b)

Fig. 3. Silhouette representation: (a) Boundary extraction (b) Distance between the centroid and each border pixel [3]

By calculating the distance from the centroid for each border pixel thanks to the expression (4), the data contained in the silhouette are converted into a 1D signal.

$$d_i = \sqrt{(x_i - x_c)^2 + (y_i - y} \qquad (4) \qquad [3]$$

3) *Signal Normalization*

The signal is normalized by the vertical distance traced in the figure n°3 a) in order to remove the impact of the scale. The fact this distance is quite constant during the walk, the person can be captured at 5m or 8m, only the relative position of boundary points is needed.

In addition to that, the size of the signal is fixed to 360 points by interpolation so that the signals have the same size for the similarity measures.

D. Eigenspace Computation

The purpose of this step is to generate a new space more adapted to analyze the similarity between the motions. Based on the gait of each person, the system computes a new space. Assuming the system has to be able to recognize L individuals. Each person is represented in the program by a sequence of distance signals. The algorithm uses as training set all the distance signals: $[D_{11}, D_{12}, \ldots, D_{1N}, D_{21}, D_{22, \ldots,} D_{LN}]$ where D_{ij} is the j^{th} distance signal of the person i and N the number of distance signal per person.

The program computes the mean μ of this training set and also the covariance matrix ∑:

$$\mu = \frac{1}{N \times i} \cdot \sum_{i=1}^{L} \sum_{j=1}^{N} D_{ij} \qquad (5)$$

$$\Sigma = \frac{1}{N \times L} \cdot \sum_{i=1}^{L} \sum_{j=1}^{N} (D_{ij} - \mu)(D_{ij} - \mu)^T \qquad (6)$$

By calculating the eigen vectors of ∑, the programs get a projection matrix W which enables to project distance signals into a new analysis space.

$$P_{ij} = W^T \cdot D_{ij} \qquad (7)$$

E. Recognition

The training set projected in the new space thanks to (7), the system is able to recognize people. Let I_u be a distance signal from an unknown walking person. The program subtracts the mean μ from I_u and projects this result in the eigenspace to perform similarity measure (8).

$$I_t = I_u - \mu$$

$$P_u = W^T \cdot I_t$$

$$\Delta = \| P_u - P_{ij} \| \quad (8)$$

$$i \in \{1, \ldots, L\} \, j \in \{1, \ldots, N\}$$

The closest distance signal P_{ij} in the database is considered to represent the silhouette I_u. By counting the number of silhouettes which are similar to those in the database, the program plots a histogram with the person on the abscissa. The maximum bar corresponds to the identified person. As we can see in figure n° 5 for example, the person number 4 is identified.

IV. Experiments And Analysis

Experiments have been done in order to verify the effectiveness of the program. The following parts describe the details of the tests.

A. Database Creation

1) Equipment Used

The video for the training and tests stages have been captured with a camera OLYMPUS SP-51OUZ fixed on a tripod in SHQ mode (30 fps and 640x480). Then, the resolution of each video has been also reduced by 2 in order to save computation time.

2) Database Structure

The database contains videos of 20 individuals which are walking along a wall. These people have been filmed two times: one passage for the training stage and another one for the test. These persons have been filmed in lateral view at 6m as we can see in figure n°5 (a).

The database contains 40 videos: 20 for the training stage and 20 for the tests.

The training videos have been cut so that they contain at least 1 stride starting when the individuals stand with their feet together. These learning videos are 61 frames long. The test sequences have not been cut.

3) Encountered Difficulties

The quality of the input data must be good for discriminating people. A lot of attempts have been made to create this database. Indeed, we have made 4 attempts before obtaining and learning all the required conditions for the identification. As a result, the system works only if we respect the following operating conditions:

- The contrast between the human and the background needs to be high. Indeed, identification has failed 1 time in our experiment due to the color of the clothes which is too similar to the background.
- The light where the video are taken must be as constant as possible. Indeed, we noticed that when the sun light changes suddenly, it creates errors in the extraction of the silhouette. Perhaps, even if the light is constant, there is a problem with the shadow of the person. Indeed, this shadow is considered by the system as a moving object too.
- The weather needs to be as good as possible. The wind is a source of perturbation which introduces noise in the segmentation.

B. From Video Processing to Identification

The algorithm is implemented in Matlab code. The program starts by converting the videos in YCbCr in order to work on the Y component. The passage of an object in front of the camera is more visible on the Y channel.

Then, the background is estimated with the technique described in part III B 3.The result of this operation is illustrated in the figure n°4.

Fig. 4 Result of median background estimation over 60 frames

Using background subtraction techniques described in part III B 4, the algorithm extracts the human silhouette. Then, the system converts these data in 1D like it is explained in part III C 2. Repeating these operations for each silhouette of each individual, the program generates a new analysis space and registers the motion of people in the database.

By measuring the similarity between the extracted silhouettes and those of each person in the database, the system generates a histogram representing the number of extracted silhouettes corresponding to each individual registered in the database. The system considers that the person in the database having the highest bar in this histogram is the walking person in the test video (figure n°5).

C. Advantages and Weaknesses of the Identification System

1) Advantages

The “security” of the identification in the histogram is quite good (see figure n°5 or appendix B for more details). Indeed, the maximum bar which corresponds to the identified person is in average at least twice superior to the second maximum. Moreover the silhouette recognition accuracy is 48% in average.

The capacity to be less sensitive to the speed of the walk is a good point. Indeed, people can walk at different speed keeping the same posture if we admit that the gait of a person is the same between 3 and 6 km/h approximately. The gait changes when individuals exceed a certain speed.

Most of people in the database wear clothes which are quite similar. The gait is the only source of data to do the identification. This experiment confirms that gait is a rich source of information which enables identification.

The computation time for the training stage is 112 sec for 20 persons which correspond to 5,6 sec per person with an 2,4GHz Intel Core 2 Duo. The recognition of a person in a video during 5 sec in average is 14 sec. By knowing that performances of processor are improved each year and also that the matlab code used here is not optimized, this algorithm proposed might be used for real time application if the code is translate in C code.

2) *Weaknesses*

Sometimes, the system cannot identify people. Indeed, the recognition has failed one time in our experiment because the walking person was wearing a grey jacket and the color of the background was light brown. As a result, any segmentation is possible: the system is blind.

The quality of video segmentation is a key step of the algorithm because it is the only source of data which enables the recognition. Although the system recognizes the person with the segmentation illustrated in figure n°6, it could have corrupted the vision of the system with a failed identification. Indeed, if the program does several bad segmentations, the system will have difficulties to distinguish people.

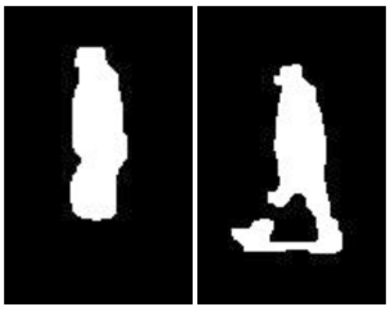


Fig. 6 Bad segmentation of a video for person 11: feet disappear

Another weakness is the direction of the walk. The implemented system does not enable the recognition of people in both directions. This problem can be solved by doing symmetry of the silhouette in function of the direction of the motion in the learning and test video.

People in the database wear the same clothes for the training stage and the test. If people are filmed with different clothes, the recognition rate will decrease. Indeed, the clothing characteristics (slim, large) modify the silhouette and alter the vision of the system.

The number of people in the database is quite small here. The recognition rate of the system will decrease at a certain number of people. Indeed, the probability to do errors in the discrimination of silhouettes increases when the system has to be able to recognize a large set of people.

D. Possible Application Domains

The system has been tested with a database containing 20 persons and the accuracy is 95%. However, these good results are obtained only if we respect some conditions which are described in IV A 3.

The fact the system is sensitive to the contrast between the clothes and the background, it can be used in environments where people have the same clothes: companies and army. This system can be used for automatic access control. For example, we can imagine that people are filmed in a corridor and the gate opens itself if the individual has the authorization. The advantage is that the person has not to put his hand on a fingerprint sensor.

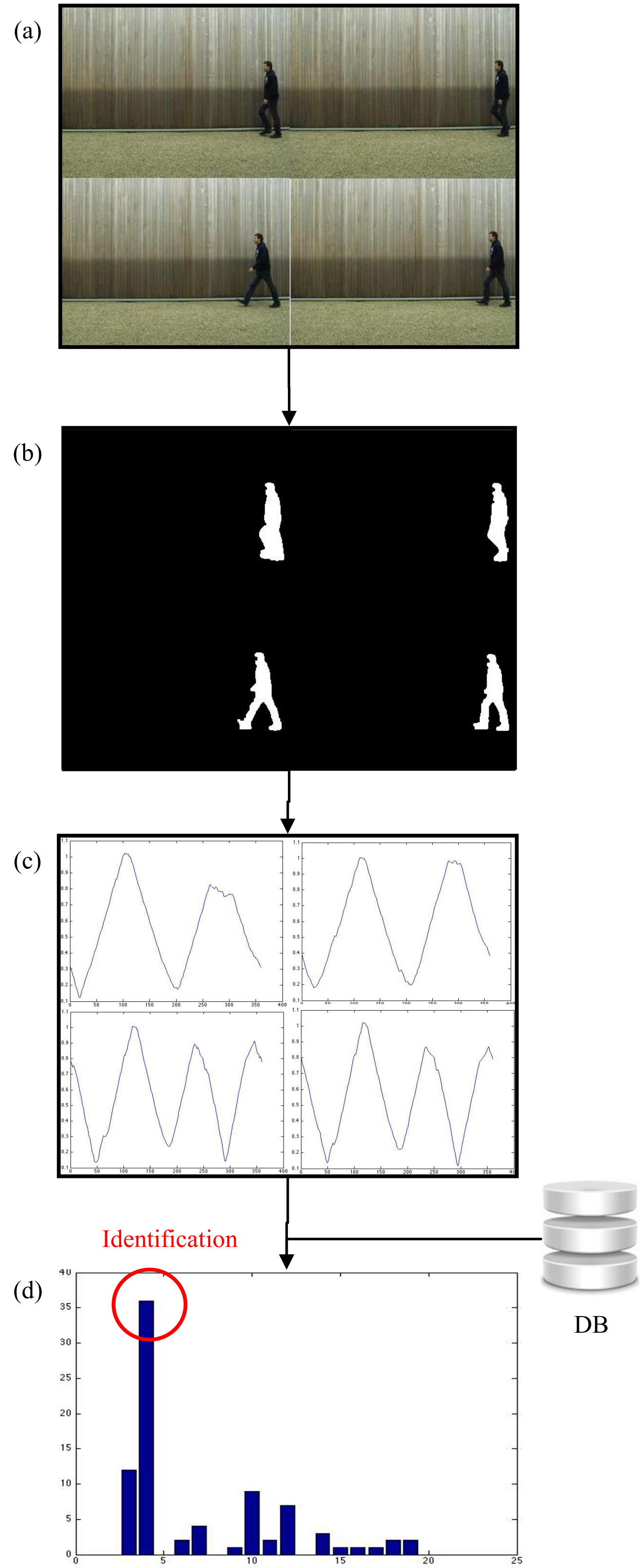


Fig. 5 Identification process: (a) 4 frames of a test sequence. (b) Respective extracted silhouettes for the 4 frames. (c) Distance signal corresponding to the silhouettes. (d) Histogram representing the number of silhouette found in the test sequence for each person.

V. Conclusions

Analyzing the gait of people to do identification is a complex problem because it requires to gather a certain number of operating conditions. This paper confirms that gait is a rich feature for biometrics. Indeed the system presented in this paper is able to identify people with an accuracy of 95%.

The algorithm proposed here is efficient and is easy to implement. Its low computational cost and its speed independence are two significant advantages. The results obtained with this algorithm are very encouraging.

Although, the system is working, the next step of this work is to know its exact limits:

- What is the maximum number of people we can include in the database keeping a reasonable accuracy?
- Here, we keep all the eigen vectors for the projection matrix. How much can we reduce the dimension of the input data keeping a good precision?

The method in this paper does not focus on the identification in the both walking direction. A new improvement could be the use of symmetry between silhouettes to add this feature to the system.

A future experiment with another segmentation technique to extract the silhouette could be the use of a laser camera working like a sonar. The combination of the segmentation done from the laser and color camera could perhaps give a better silhouette extraction.

Acknowledgment

I would like to express my gratitude to all those who gave me the possibility to achieve my master project. I want to thank my supervisor, Paul Whelan, Professor of computer vision whose stimulating suggestions and encouragement helped me in my research. I thank also Dr Ovidiu Ghita for his help and guidance.

I am especially grateful to my friends and family who agreed to participate in my project and take time to be filmed. I also thank the students of the French master with whom the discussions were enriching and motivating.

I would like to give my special thanks to my girlfriend whose support helped me in this work.

References

[1] C.-C. Lien, C.-C. Tien and J. M. Shih “Human Gait Recognition for Arbitrary View Angles”, *Innovative Computing, Information and Control*, 2007.pp.303, Sep. 2007.

[2] M. S. Nixon, T. Tan and R. Chellappa, “Human Identification Based on Gait”, *Springer Science and Business*, pp. 185-203, 2006.

[3] L. Wang, T. Tan, H. Ning and W. Hu “Silhouette Analysis-Based Gait Recognition for Human Identification”, *Pattern Analysis and Machine Intelligence, IEEE Transactions* on, pp. 1505-1518, Dec 2003.

[4] C. BenAbdelkader, R. Cutler and L. Davis, “Stride and Cadence as a Biometric in Automatic Person Identification and Verification”, *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, 2002.

[5] S. V. Stevenage, M. S. Nixon and K. Vince, “Visual Analysis of Gait as a Cue to Identity”, *Applied Cognitive Psychology* vol 13, pp. 513-526, 1999.

[6] P. S. Huang, C J. Harris and M. S. Nixon, “Human Gait Recognition in Canonical Space Using Temporal Templates”, *IEEE Proc.Vision Image and Signal Processing Conf.*, vol. 146, n°2, pp. 93-100, 1999.

[7] S. Sulaiman, A. Hussain, N. Tahir, S.A. Samad and M.M. Mustafa, “Human Silhouette Extraction Using Background Modeling and Substraction Techniques”, *Information Technology Journal*, 7(1), 155-159, 2008.

[8] S-C. S. Cheung and C. Kamath, “Robust Techniques for Background Subtraction in Urban Traffic Video” *EURASIP Journal on Applied Signal Processing*, pp. 2330-2340, 2005.

[9] Paul F Whelan (2011), “VSG Image Processing and Analysis (VSG IPA) Toolbox”, Technical Report, Centre for Image Processing & Analysis, Dublin City University

[10] Paul F Whelan (2011), “VSG Image Processing and Analysis (VSG IPA) Matlab Toolbox - Software””, http://www.cipa.dcu.ie/code.html, Centre for Image Processing & Analysis, Dublin City University.

[11] Paul F. Whelan and D. Molloy (2000), "Machine Vision Algorithms in Java: Techniques and Implementation", Springer (London), 298 Pages. ISBN 1-85233-218-2.

[12] Paul F Whelan (2011), "EE544: Computer Vision Course Notes”, Centre for Image Processing & Analysis, Dublin City University.

[13] Paul F Whelan (2011), "EE425/EE453: Image Processing and Analysis Course Notes”, Centre for Image Processing & Analysis, Dublin City University.